The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. A computer-implemented method for selectively accessing a document during a current crawl of a server computer, the document being identified by a document address specification, the document having been retrieved during a previous crawl, the method comprising:

determining whether to access the document during the current crawl with the aid of a statistical model; and

accessing the document if the determination produces an instruction indicative that the document at the document address specification should be accessed during the current crawl.

2. The method of Claim 1, wherein determining whether to access the document further comprises computing a probability that the document has changed since the document was retrieved during the previous crawl.

3. The method of Claim 2, wherein computing the probability that the document has changed further comprises:

selecting an active probability indicative of a proportion of documents in a plurality of documents that are changing at various change rates, the plurality of documents including the document;

training the active probability to reflect an experience with the document during a plurality of previous crawls; and

using the trained active probability to compute the probability that the document has changed.

4. The method of Claim 3, further comprising:

selecting the probability that the document has changed from the previous crawl as the active probability in the current crawl; and

repeating the method of Claim 3 for the current crawl.

5. The method of Claim 3, wherein training the active probability includes multiplying the active probability indicative of a change in the document by a training probability calculated using a statistical model.

6. The method of Claim 1, wherein the statistical model further comprises:

training a document probability distribution corresponding to the document address specification to reflect an experience with the document during a plurality of previous crawls, the document probability distribution including a plurality of probabilities;

determining from the document probability distribution a probability that the document has changed; and

making a determination of whether to access the document in a current crawl based on the probability that the document has changed.

7. The method of Claim 6, further comprising:

calculating, based on the experience with the document during a plurality of previous crawls, a discrete random variable distribution that includes a plurality of training probabilities;

multiplying each probability in the document probability distribution by a corresponding training probability from the discrete random variable distribution.

8. The method of Claim 7, wherein the training probabilities are calculated using a Poisson process, the Poisson process including a Poisson equation (e^(-r*dt)) and a complementary Poisson equation (1-e^(-r*dt)).

9. The method of Claim 8, wherein the experience with the document during the plurality of previous crawls is derived from historical information associated with the document address specification.

10. A computer-readable medium having computer-executable instructions for retrieving one document in a plurality of documents from a remote server, which when executed comprise:

maintaining historical information associated with changes to the one document at the remote server;

initiating a crawl procedure for retrieving particular documents in the plurality of documents; and

determining whether to access the one document from the remote server based on an analysis of the historical information associated with the changes to the one document at the remote server.

11. The computer-readable medium of Claim 10, further comprising:

if the determination to access the one document is positive, identifying the one document for retrieval during the crawl procedure; and

attempting to retrieve all documents identified for retrieval during the crawl procedure.

12. The computer-readable medium of Claim 10, wherein determining whether to retrieve the document further comprises:

computing a probability that the one document has changed since the one document was last retrieved from the remote server.

13. The computer-readable medium of Claim 12, wherein computing the probability that the one document has changed further comprises:

beginning with a probability that a pre-defined proportion of documents in the plurality of documents has changed, training the probability that the pre-defined proportion of documents has changed using the historical information associated with the one document to achieve the probability that the one document has changed.

14. The computer-readable medium of Claim 12, further comprising making a random decision to retrieve the one document wherein the random decision is biased by the probability that the one document has changed.

15. The computer-readable medium of Claim 14, wherein the random decision is further biased by a synchronization level configured to influence the random decision based on a predetermined degree of tolerance for not retrieving the one document if the document is likely to have changed.

16. The computer-readable medium of Claim 14, wherein the random decision is made by a software routine adapted to simulate a flip of a coin.

17. The computer-readable medium of Claim 10, wherein:

the historical information associated with changes to the one document includes a time stamp for the one document, the time stamp being indicative of a last time that the one document was modified when the one document was last retrieved from the remote server; and

wherein the analysis includes a comparison of the time stamp included in the historical information with another time stamp associated with the one document stored on the remote server.

18. The computer-readable medium of Claim 17, further comprising:

if the time stamp included in the historical information does not match the other time stamp associated with the one document stored on the remote server, identifying the one document for retrieval during the crawl procedure.

19. The computer-readable medium of Claim 10, wherein:

the historical information associated with changes to the one document includes a hash value associated with the one document, the hash value being a representation of the one document; and

wherein the analysis includes a comparison of the hash value included in the historical information with another hash value calculated from information retrieved from the one document stored on the remote server.

20. The computer-readable medium of Claim 19, if the hash value included in the historical information does not match the other hash value associated with the one document stored on the remote server, identifying the one document for retrieval during the crawl procedure.